Evaluating 11 Years of Quantitative Precipitation Forecast Performance for Extreme Events

Ellen Sukovich[1,2], Marty Ralph[2], David Novak[3], and Faye Barthold[3,4]

[1]*Cooperative Institute for Research in the Environmental Sciences, University of Colorado*
[2]*NOAA/OAR /Earth System Research Laboratory*
[3]*NOAA/NWS/NCEP Hydrometeorological Prediction Center*
[4]*I.M. Systems Group, Inc.*

Extreme precipitation events (i.e., those events associated with the tail end of the precipitation probability distribution) are impactful events that can cause loss of life, damage to property, and significant disruption to local, regional, and even national economies. There are many communities, such as water resources management, industry, agriculture, transportation, government, and emergency management, which require accurate forecasts of these extreme events for decision-making, preparation, and management; however, accurately forecasting such events remains one of meteorology's most difficult challenges. Since verification provides both a way to measure improvement in quantitative precipitation forecasts (QPF) and a method by which forecast errors can be identified, the Hydrometeorology Testbed (HMT) has identified QPF verification as an integral component to improving forecasts of extreme precipitation events.

This study examines QPF performance for extreme events over 11 years (January 1, 2001 through December 31, 2011) to evaluate trends in forecasting performance necessary to quantify and improve the timing, location, and amount of predicted precipitation. Using HPC's 32-km gridded QPF data and per Ralph et al. (2010), regional extreme precipitation thresholds were determined by calculating the top 1% and top 0.1% of events for each River Forecast Center (RFC) region. Using the Developmental Testbed's (DTC) Model Evaluation Tools (MET) verification package, five primary measures were used to analyze QPF performance: probability of detection (POD), false alarm ratio (FAR), critical success index (CSI), mean absolute error (MAE), and bias. The results of these metrics were compared to the current NOAA Government Performance and Results Act (GPRA) precipitation threshold (≥ 1.0 in 24 h^{-1}), which is more representative of moderate precipitation events, to determine a baseline performance.

Preliminary results from this study indicate that 32-km extreme QPF has improved over the last 11 years, even though the yearly CSI values (i.e., threat scores) of the baselined extreme precipitation are approximately half of the GPRA threat scores. In addition, extreme QPF threat scores appear to be improving slightly faster ($\sim$10-15%) than the GPRA threat scores ($\sim$9%) between 2001 and 2011. Further examination has also shown that extreme precipitation amounts tend to be consistently under predicted. A key challenge of this verification work is the smaller sample size of the extreme events, which tend to occur less frequently, and the subsequent statistical robustness of the findings. Finally, results from this work will be communicated to the operational forecast community to begin the process to better monitor the performance of extreme QPFs.